



Bias in Self-reported Voting and How it Distorts Turnout Models
Disentangling Nonresponse Bias and Overreporting Among Danish Voters

Dahlgaard, Jens Olav; Hansen, Jonas Hedegaard; Hansen, Kasper M.; Bhatti, Yosef

Published in:
Political Analysis

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY-NC-SA](#)

Citation for published version (APA):
Dahlgaard, J. O., Hansen, J. H., Hansen, K. M., & Bhatti, Y. (2019). Bias in Self-reported Voting and How it Distorts Turnout Models: Disentangling Nonresponse Bias and Overreporting Among Danish Voters. *Political Analysis*, 27(4), 590-598.

Bias in Self-reported Voting and How it Distorts Turnout Models: Disentangling Nonresponse Bias and Overreporting Among Danish Voters

Jens Olav Dahlgaard¹, Jonas Hedegaard Hansen²,
Kasper M. Hansen³ and Yosef Bhatti⁴

¹ Department of International Economics, Government and Business, Copenhagen Business School, Frederiksberg, Denmark. Email: jod.egb@cbs.dk

² Independent Researcher, Copenhagen, Denmark. Email: hedegaardhansen@gmail.com

³ Department of Political Science, University of Copenhagen, Copenhagen, Denmark. Email: kmh@ifs.ku.dk; URL: www.kaspermhansen.eu

⁴ VIVE—The Danish Centre of Applied Social Science, Copenhagen, Denmark. Email: yobh@vive.dk

Abstract

Most nonexperimental studies of voter turnout rely on survey data. However, surveys overestimate turnout because of (1) nonresponse bias and (2) overreporting. We investigate this possibility using a rich dataset of Danish voters, which includes validated turnout indicators from administrative data for both respondents and nonrespondents, as well as respondents' self-reported voting from the Danish National Election Studies. We show that both nonresponse bias and overreporting contribute significantly to overestimations of turnout. Further, we use covariates from the administrative data available for both respondents and nonrespondents to demonstrate that both factors also significantly bias the predictors of turnout. In our case, we find that nonresponse bias and overreporting masks a gender gap of two and a half percentage points in women's favor as well as a gap of 25 percentage points in ethnic Danes' favor compared with Danes of immigrant heritage.

Keywords: measurement error, administrative data, validated turnout, voter turnout

Voter turnout is the modal form of political participation in established democracies, and scholars have long paid attention to who votes in elections by comparing turnout between subgroups or by predicting turnout based on background covariates (Tingsten 1937; Wolfinger and Rosenstone 1980; Rosenstone and Hansen 1993). We show that different measures of turnout can lead to drastically different conclusions and methodological choices have important implications for the inferences we make about voters and nonvoters.

Broadly speaking, there are two ways to measure turnout at the individual level. Either respondents in a survey are asked to self-report whether they voted, or official voter records are used to validate who voted. A recent meta-analysis by Smets and van Ham (2013, 346) shows that within the top 10 political science journals, only 11% of articles employing turnout models used validated data.¹

Despite the prevalence of surveys in the study of voter turnout, it has long been recognized that surveys overestimate the turnout rate (e.g. Clausen 1968; Traugott and Katosh 1979; Silver,

Authors' note: This work was supported by the Danish Council for Independent Research (grant no. 12-124983). The paper builds on work previously published in Danish by the same authors (Bhatti et al. 2017). We are grateful to Florian Foos, Mogens K. Justesen, Michael Goldfien, and seminar participants at Copenhagen Business School for comments and suggestions on previous versions of this paper. The authors are listed alphabetically by their first name. The Danish National Election Studies is freely available through the Danish National Archives. The Carlsberg Foundation financed the data collection for DNES 2015 (grant no. CF14-0137, Hansen and Stubager 2016). Replication materials for this paper are posted to the Dataverse of Political Analysis (Dahlgaard 2018). The microdata linked to the administrative data can unfortunately not be uploaded due to privacy concerns imposed by the data provider, Statistics Denmark.

1 In the Get-Out-The-Vote literature, the picture is different as the norm is to use validated turnout.

Political Analysis (2019)
vol. 27:590–598
DOI: 10.1017/pan.2019.9

Published
13 March 2019

Corresponding author
Jens Olav Dahlgaard

Edited by
Jeff Gill

© The Author(s) 2019. Published by Cambridge University Press on behalf of the Society for Political Methodology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

Anderson, and Abramson 1986; Granberg and Holmberg 1991; Bernstein, Chadha, and Montjoy 2001; Karp and Brockington 2005; Jackman and Spahn 2014). Two factors contribute to this overestimation. First, nonvoters may be less likely to participate in surveys. We will call this *nonresponse bias*. Second, survey participants misreport their voting behavior, typically saying that they voted when they did not. We will call this *overreporting*.

We contribute to the literature on voter turnout by analyzing a rich dataset of Danish voters. The dataset includes validated turnout, self-reported turnout, and background covariates for respondents, as well as validated turnout and background covariates for nonrespondents. Validated turnout and covariate information come from administrative data, and self-reported turnout comes from the Danish National Election Studies (DNES) survey.

We first demonstrate that for Danish voters both nonresponse bias and overreporting contribute to the overestimation of turnout in surveys. We then estimate three predictive models: one with all of the voters in the original sampling frame for whom we have validated turnout; one with validated turnout for only survey respondents; and one with self-reported turnout from the survey. In all three models, we use highly reliable covariates from the administrative data. We compare the models to a baseline model for the population.

The relationship between turnout and age, education, and ethnicity is weaker among the survey participants due to nonresponse bias. It becomes even weaker when we use self-reported turnout due to overreporting among respondents. When we use a self-reported measure of turnout from a survey, as most published research does, both a turnout gap of two and a half percentage points between men and women and of 25 percentage points between native Danes and non-natives disappear. That is, turnout measures that rely on self-reported data may mask important covariate relationships. We conclude that researchers should use validated turnout data when possible. If self-reported voter turnout is the only data available, researchers should use question wordings that reduce overreporting, but also be aware that both nonresponse and overreporting can bias their results.

1 Overreporting and Nonresponse Bias in Surveys

Why could overreporting and nonresponse bias in surveys affect the characterization of voters and nonvoters? First, voters may be more likely than nonvoters to participate in postelection surveys (Katosh and Traugott 1981). Thus, responding to the surveys may correlate with propensity to vote. Likewise, once respondents enter surveys, overreporting is not randomly distributed among responding nonvoters. Rather, several studies have shown that overreporting is correlated with predictors of turnout (e.g. Bernstein, Chadha, and Montjoy 2001; Ansolabehere and Hersh 2012; Brenner 2012).

Previous research has focused on who overreports, how to reduce overreporting, and the consequences of overreporting. Studies have focused on nonvoters, to account for who is overreporting (Silver, Anderson, and Abramson 1986; Granberg and Holmberg 1991; Brenner 2012), and techniques to reduce overreporting (Abelson, Loftus, and Greenwald 1992; Morin-Chassé *et al.* 2017). For instance, Karp and Brockington (2005) explore predictors of overreporting across five countries and find that overreporting is positively associated with turnout and positively associated with predictors of voting. Ansolabehere and Hersh (2012) show that using validated turnout leads to different estimates for predictors of turnout compared to using self-reported turnout. Among others, Belli *et al.* (1999), Belli, Moore, and VanHoewyk (2006), Hanmer, Banks, and White (2014) show that paying attention to question wording and response categories can remove some of the bias from overreporting.

Studies have also focused on nonresponse (Swaddle and Heath 1989; Sciarini and Goldberg 2016). Nonresponse typically biases the estimates of turnout, since voters are more inclined to participate in surveys. Sciarini and Goldberg (2016) find that nonresponse bias may also lead to

misestimation of the predictors of overreporting. This highlights the need to consider not only how overreporting and nonresponse lead to turnout overestimation but also how they may bias the characterization of voters and nonvoters (Ansolabehere and Hersh 2012).

2 Data and Estimation Strategy

We combine the DNES from 2015 with Danish administrative data. The 2015 election took place in June and had an overall turnout rate of 85.9%. All Danes have a unique civil registration number, which is linked to administrative data containing a range of background information including age, sex, education, and ethnicity. The DNES sampling frame was drawn from the list of civil registration numbers for all Danish voters. As the civil registration numbers are unique, they provide sufficient information for uniquely matching the entire sampling frame to the administrative data.

Everyone who meets the eligibility requirements is automatically registered to vote, which guarantees that everyone sampled for the DNES was in fact eligible to vote. The voter lists for Danish elections are created based on the administrative data and can therefore also be linked directly to the civil registration number. This means that we can uniquely link turnout to the administrative data for the entire sampling frame, and to survey responses for those who actually took the survey. Because the final voter lists are created approximately one week before each election, there are close to no voters that figure on the lists even though they have died or moved. In 2015, we collaborated with 72 out of 98 municipalities to link turnout for their citizens to the administrative data (Bhatti *et al.* 2016).² Where we linked turnout to administrative data, we have a voter record for everyone sampled for the DNES.

The DNES was carried out after the election (Hansen and Stubager 2016). It consisted of a probability sample drawn from the civil registration numbers of voters less than 76 years of age with an oversample of young voters. Of the sampled subjects, 3,145 lived in one of the 72 municipalities for which we have validated turnout. The response rate in these municipalities was 52.0%, meaning that 1,635 subjects opted to participate. The DNES records turnout with a multiple-choice question for which one response option was “did not vote”. As both respondents and nonrespondents to the survey were linked to the administrative data, we have validated turnout for the sample, which is free from both nonresponse bias and overreporting. We also have validated turnout for the respondents, which suffers from nonresponse bias but not overreporting. Finally, we have self-reported turnout from the sample, which is subject to both overreporting and nonresponse bias.

First, we show that both nonresponse bias and overreporting contribute to the inflated turnout estimates. To learn about the contribution of self-selection, we compare validated turnout between the 1,510 who did not respond to the survey and the 1,635 who did. To learn about the contribution of overreporting, we compare self-reported turnout with validated turnout among those who participated in the survey.

We rely on a similar strategy to learn whether and how nonresponse bias and overreporting biases the relationship between turnout and background covariates. Based on covariates from the administrative data, we can make three predictions. First, we can predict the probability of voting in the full sample of validated voters. Second, we can predict the probability of voting in the sample of voters who participated in the DNES. Third, we can predict the probability of reporting voting in the DNES. As the first estimate will be based on the entire probability sample, it should fall within the sample variation of the population estimate and will serve as the baseline. The second estimate will suffer from nonresponse bias. Misreporting will further bias the third estimate.

² Municipalities use either digital voter lists where voters hand in a polling card or paper voter lists where a pollworker manually marks when the voter turns out. Turnout was linked for all voters in the municipalities that relied fully on digital voter lists and 12 municipalities that relied on paper voter lists.

Table 1. Validated and self-reported turnout for DNES sample frame and respondents.

Turnout measure	Population	DNES sample	Nonrespondents	Respondents	
	Validated	Validated	Validated	Validated	Self-reported
Turnout rate (%)	86.4	86.2	77.9	93.3	96.2
N	2,810,312	3,144	1,510	1,634	1,621

Eleven respondents who answered a few questions but dropped out before the turnout question were omitted and counted as nonrespondents. Thirteen of the respondents answered “don’t know” to the turnout question. Columns 2–4 are weighted to adjust for oversampling of young voters. The population is voters younger than 76 with validated turnout.

3 Results

To account for the oversampling of young voters, we use inverse probability weights in all analyses for the sample; that is, we weight all observations with one divided by the probability of being sampled. Population results are unweighted. In the first column of Table 1, we show the turnout rate, 86.4%, for the population of voters younger than 76 with validated turnout; the population that the DNES with validated turnout was sampled from. The second column displays the turnout rate, 86.2%, for the entire sampling frame. In the third and fourth columns, we show validated turnout among nonrespondents and respondents. Evidently, the turnout rate is substantially higher due to nonresponse bias; 93.4% among respondents compared to 77.9% among nonrespondents. The proportion of nonvoters is roughly halved from 13.8% in the sampling frame to 6.6% in the sample. Self-reported turnout is even higher, at 96.2%, due to overreporting. Once again, the proportion of nonvoters is almost halved, from 6.6% to 3.8%.

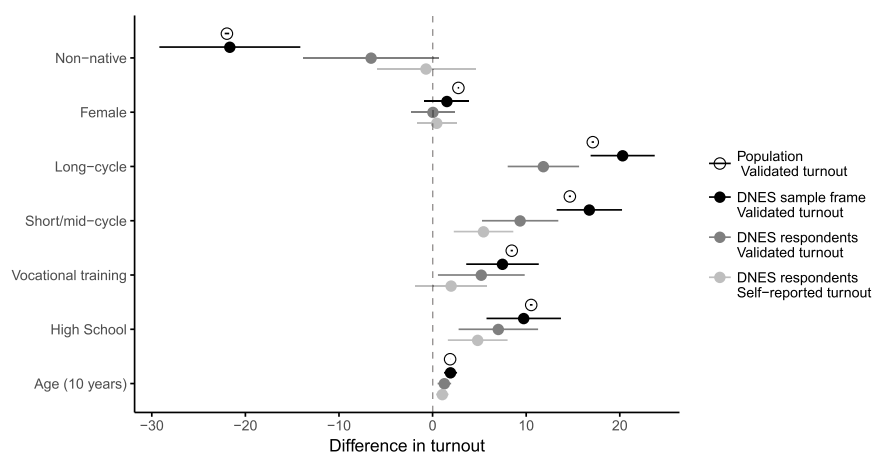
The first set of results shows that both self-selection and overreporting among respondents drives up survey estimates of voter turnout. The next question is whether nonresponse and overreporting matter for the substantive conclusions we draw about voter turnout in a predictive model. In Figure 1, we run four models using different covariates to predict turnout. We include as background covariates age in years, education (which is coded as 1 if the respondent has four or more years of education beyond high school), the subject’s sex, and an indicator for being an immigrant or immigrant descendant.³ We choose these covariates because they are widely used in turnout studies, and we know from previous research that they predict turnout for Danish voters (Smets and van Ham 2013; Bhatti *et al.* 2018).

As a starting point, we predict turnout for everyone for whom we have validated turnout. In the second model, we include everyone who was invited to participate in the DNES, and we estimate the relationship between validated turnout and background covariates. In the third model, we estimate the relationship between validated turnout and background covariates only among those who participated in DNES. Finally, in the fourth model, we use self-reported turnout for the DNES respondents with validated turnout while still using background covariates free of misreporting from the administrative data.⁴ For each model, we estimate a logistic regression and in the figure, we present average marginal predictions.

When we look at the actual relationship between the covariates and turnout in 2015, we see that older and better-educated citizens, females, and native Danes were more likely to vote. These findings align with previous research (Smets and van Ham 2013; Bhatti *et al.* 2018). As we would expect from a random sample, the point estimates in the sampling frame model are close to the

3 These are official categories created by Statistics Denmark. According to the definitions, an “Immigrant” is someone born abroad and neither of his or her parents are both Danish citizen and born in Denmark. A “descendant” is born in Denmark and neither of his or her parents are both Danish citizen and born in Denmark. We follow this definition.

4 The last model comes closest to the standard survey approach, although many surveys also rely on self-reported background covariates with additional uncertainty.



The coefficients are marginal predictions in percentage points based on a logistic regression. In the supporting information, we present the marginal predictions in table format. There is no prediction for self-reported turnout among voters with long-cycle education as they all reported voting. The analyses are weighted by the inverse probability of being sampled for the DNES. For the marginal predictions, other variables are held at their observed values, which is the most representative approach (Hanmer and Ozan Kalkan 2013).

Figure 1. Predicted turnout for population, sample frame, and respondents.

population model. The population estimates are even included in all 95% confidence intervals for the sampling frame.

In model 3 (dark gray), we consider only validated turnout for the survey respondents, which tells us what happens when we introduce nonresponse bias. We would still believe that older, better educated, and native Danes were more likely to vote. However, the estimates are substantially smaller in magnitude and the difference in turnout between men and women disappears. In the fourth model (light gray), when we use the survey measure of turnout, which is subject to both overreporting and nonresponse, our conclusions become even weaker. Older and better-educated voters are still more likely to report voting, but the magnitude of the difference is now even smaller. The non-native Danes' turnout is comparable to that of native Danes. A gender difference of more than 2 percentage points has also disappeared.

3.1 Predicting Survey Participation

To provide a clearer response as to why nonresponse biases the correlates of voting, we show in Table 2 how the covariates that we use to predict voter turnout also predict survey participation. In the table, we estimate a model with survey response as the dependent variable for both the entire sampling frame and only the part for which we have validated turnout. If voters are more prone to take the survey, we should see that what predicts turnout also predicts survey participation, effectively leading to truncation on the dependent variable. We see exactly this pattern: age and education are strong predictors of participation, just as they are of voter turnout. Older respondents and respondents with more education are more likely to take the survey. Non-natives are less likely to take the survey. It does not appear that either sex is more likely to take the survey.

We can also compare coefficients from the two models to see whether our results are biased by only some municipalities reporting turnout. The two models provide qualitatively similar results. If the results had differed, it would have caused us to worry how the results were affected by municipalities opting into the study.

3.2 Sensitivity to Missing Turnout

It is worth taking a closer look at whether missing data could have an impact on our conclusions. In the top panel of Table 3, we show weighted descriptive statistics for the survey participants in columns 1–3. In columns 4–6, we show equivalent statistics for all Danish citizen below the age

Table 2. Predicted probability of survey participation.

<i>Dependent variable</i>	Sampling frame with validated turnout	Full sampling frame
	Survey participation	Survey participation
Age	0.35 (0.06)	0.35 (0.05)
Education (Baseline = Elementary school)		
High school	7.94 (3.08)	10.35 (2.79)
Vocational training	4.36 (2.48)	7.66 (2.11)
Short/mid-cycle (Up to a Bachelor's)	18.72 (2.67)	21.93 (2.31)
Long-cycle (A Master's or more)	23.06 (3.62)	25.52 (3.20)
Female	−2.79 (1.85)	−1.72 (1.61)
Non-native Dane	−15.16 (4.25)	−12.81 (3.95)
N	3,107	4,061
McFadden's R ²	0.03	0.04

The coefficients are marginal predictions in percentage points based on a logistic regression. Standard errors in parenthesis.

of 76. In the middle panel, we show statistics only for citizens and survey respondents without validated turnout. In the bottom panel, we show descriptive statistics for citizens and survey respondents in all municipalities where we know turnout.

We see that the survey respondents for whom we have validated turnout tend to be younger and more likely to have a high school education or a long-cycle education, while they are less likely to have received vocational training. The variations reflect differences in the populations from which the sample was drawn. When we concentrate on the populations, we see similar differences between the populations in municipalities with validated turnout and the populations in the municipalities without. Importantly, there seems to be no huge discrepancies between the populations of participating and nonparticipating municipalities.

In the supporting information, we also look at the geographic distribution of the nonparticipating municipalities. Furthermore, we run a sensitivity test where we substitute validated turnout in 2015 with validated turnout from a 2013 election where we have full population validated turnout but no polling data. If we assume that self-selection would have been the same in 2013, we can estimate the consequence of nonresponse bias for the entire population in that election and compare it to the consequence among only respondents in participating municipalities. Given this assumption, the consequence of nonresponse bias in 2013 would have been qualitatively the same for the predictive model when using full population data. Finally, we show that a model with self-reported turnout for the entire sample gives similar results as the model using only the validated turnout sample.⁵

⁵ In the supporting information, we also show additional descriptive statistics, and we present models predicting turnout among nonrespondents and a model where we focus on overreporting among nonvoting respondents. We leave additional description of these results for the supporting information.

Table 3. Descriptive statistics conditional on having validated turnout.

	All survey respondents			Danish voters (age <76)		
	Mean	Sd	N	Mean	Sd	N
Age (years)	48.5	15.6	2,172	46.7	16.2	3,747,832
Elementary School (%)	22.0	41.4	2,146	27.2	44.5	3,698,696
High School (%)	8.2	27.5	2,146	9.2	28.9	3,698,696
Vocational training (%)	34.0	47.4	2,146	34.7	47.6	3,698,696
Short/mid-cycle education (%)	26.0	43.9	2,146	20.9	40.7	3,698,696
Long-cycle education (%)	9.9	29.8	2,146	8.0	27.1	3,698,696
Female (%)	49.6	50.0	2,172	49.9	50.0	3,747,832
Native Dane (%)	96.7	17.8	2,172	95.5	20.8	3,747,832

	Respondents without validated turnout			Danish voters (age <76) without validated turnout		
	Mean	Sd	N	Mean	Sd	N
Age (years)	49.5	14.4	538	48.8	15.5	937,731
Elementary School (%)	22.4	41.7	536	29.3	45.5	927,068
High School (%)	5.0	21.7	536	5.9	23.5	927,068
Vocational training (%)	40.0	49.0	536	39.8	49.0	927,068
Short/mid-cycle education (%)	25.5	43.6	536	19.8	39.8	927,068
Long-cycle education (%)	7.2	25.9	536	5.3	22.3	927,068
Female (%)	52.6	50.0	538	49.2	50.0	937,731
Native Dane (%)	97.3	16.1	538	97.4	16.0	937,731

	Respondents with validated turnout			Danish voters (<age 76) with validated turnout		
	Mean	Sd	N	Mean	Sd	N
Age (years)	48.1	16.0	1,634	46.0	16.4	2,810,101
Elementary School (%)	21.8	41.3	1,610	26.5	44.1	2,771,628
High School (%)	9.4	29.2	1,610	10.3	30.4	2,771,628
Vocational training (%)	31.8	46.6	1,610	33.0	47.0	2,771,628
Short/mid-cycle education (%)	26.2	44.0	1,610	21.3	40.9	2,771,628
Long-cycle education (%)	10.8	31.0	1,610	8.9	28.5	2,771,628
Female (%)	48.6	50.0	1,634	50.1	50.0	2,810,101
Native Dane (%)	96.5	18.4	1,634	94.8	22.1	2,810,101

Except for a very small proportion of disenfranchised voters, the population of voters and Danish citizen are almost identical. Only voters younger than 76 was sampled for the DNES.

4 Implications

Our findings suggest that self-reported voter turnout surveys suffer from overreporting and nonresponse, which lead to upwardly biased estimates of turnout. Further, these biases can mask relationships between turnout and key covariates. In our study, using self-reported turnout instead of validated turnout would lead to the incorrect conclusion that men are about as likely to vote as women. Similarly, using self-reported turnout data instead of validated turnout would mask a substantial gap in electoral participation between native Danes and Danes of immigrant background. It goes without saying that the two approaches to measuring carry different policy implications. Survey weights could mitigate the bias created by nonresponse. We bracket that discussion, and simply reiterate that nonresponse and overreporting bias turnout models based on self-reported voting data.

While different question wording and response categories could reduce the amount of overreporting (Belli *et al.* 1999; Belli, Moore, and VanHoewyk 2006; Hanmer, Banks, and White 2014), our paper shows that even using turnout measured with little or no overreporting, for instance validated turnout, can still lead to badly biased conclusions in descriptive turnout studies. Voter turnout is an important and widely studied topic. Our purpose is not to dissuade researchers from doing survey-based turnout studies. We simply point out that researchers should think about ways to acquire high quality data that allows for the replication of established findings that may have been based on survey data, something which is not the norm in published studies (Smets and van Ham 2013). In addition to thinking about how to measure turnout in a way that is void of overreporting, efforts should also include means to reduce nonresponse bias.

Just as importantly, overreporting and nonresponse bias impact not only the turnout level, but also models predicting turnout. We should not automatically accept null findings if they are based on data that may not have the necessary quality to merit the conclusions drawn. From our point of view, researchers should invest in collecting validated turnout and administrative data as a supplement to traditional survey-based studies.

References

- Abelson, R. P., E. F. Loftus, and A. G. Greenwald. 1992. "Attempts to Improve the Accuracy of Self-reports of Voting." In *Questions about Questions*, edited by J. M. Tanur, 138–153. New York: Russell Sage Foundation.
- Ansolabehere, S., and E. Hersch. 2012. "Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate." *Political Analysis* 20(4):437–459.
- Belli, R. F., S. E. Moore, and J. Vanhoewyk. 2006. "An Experimental Comparison of Question Forms Used to Reduce Vote Overreporting." *Electoral Studies* 25(4):751–759.
- Belli, R. F., M. W. Traugott, M. Young, and K. A. McGonagle. 1999. "Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring." *The Public Opinion Quarterly* 63(1):90–108.
- Bernstein, R., A. Chadha, and R. Montjoy. 2001. "Overreporting Voting: Why it Happens and Why it Matters." *Public Opinion Quarterly* 65(1):22–44.
- Bhatti, Y., J. O. Dahlgaard, J. H. Hansen, and K. M. Hansen. 2016. "Valgdeltagelsen og vælgerne til folketingsvalget 2015." Technical report, University of Copenhagen, Department of Political Science: CVAP Working Papers Series 1/2016.
- Bhatti, Y., J. O. Dahlgaard, J. H. Hansen, and K. M. Hansen. 2017. "Valgdeltagelsen." In *Oprør Fra Udkanten*, edited by K. M. Hansen and R. Stubager, 131–150. Copenhagen: Djøf Forlag.
- Bhatti, Y., J. O. Dahlgaard, J. H. Hansen, and K. M. Hansen. 2018. "Core and Peripheral Voters: Predictors of Turnout Across Three Types of Elections." *Political Studies*, <https://doi.org/10.1177/0032321718766246>.
- Brenner, P. S. 2012. "Overreporting of Voting Participation as a Function of Identity." *The Social Science Journal* 49(4):421–429.
- Clausen, A. R. 1968. "Response Validity: Vote Report." *The Public Opinion Quarterly* 32(4):588–606.
- Dahlgaard, J. O. 2018. "Replication Data for: Bias in Self-reported Voting and How it Distorts Turnout Models: Disentangling Nonresponse bias and Overreporting among Danish Voters." <https://doi.org/10.7910/DVN/PNPR8C>, Harvard Dataverse, V1, UNF:6:iktu1UAR6U8xX+JHQ4sk9A==[fileUNF].
- Granberg, D., and S. Holmberg. 1991. "Self-reported Turnout and Voter Validation." *American Journal of Political Science* 35(2):448–459.
- Hanmer, M. J., A. J. Banks, and I. K. White. 2014. "Experiments to Reduce the Over-reporting of Voting: A Pipeline to the Truth." *Political Analysis* 22(1):130–141.
- Hanmer, M. J., and K. Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." *American Journal of Political Science* 57(1):263–277.
- Hansen, K. M., and R. Stubager. 2016. "The danish national election study 2015." Technical report, University of Copenhagen, Department of Political Science: CVAP Working Papers Series 2/2016.
- Jackman, S., and B. T. Spahn. 2014. "Why Does the American National Election Study Overestimate Voter Turnout? In *Political Methodology Meetings*, University of Georgia.
- Karp, J. A., and D. Brockington. 2005. "Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries." *Journal of Politics* 67(3):825–840.
- Katosh, J. P., and M. W. Traugott. 1981. "The Consequences of Validated and Self-reported Voting Measures." *Public Opinion Quarterly* 45(4):519–535.

- Morin-Chassé, A., D. Bol, L. B. Stephenson, and S. L. St-Vincent. 2017. "How to Survey about Electoral Turnout? The Efficacy of the Face-saving Response Items in 19 Different Contexts." *Political Science Research and Methods* 5(3):575–584.
- Rosenstone, S., and J. M. Hansen. 1993. *Mobilization, Participation and Democracy in America*. MacMillan Publishing.
- Sciarini, P., and A. C. Goldberg. 2016. "Turnout bias in Postelection Surveys: Political Involvement, Survey Participation, and Vote Overreporting." *Journal of Survey Statistics and Methodology* 4(1):110–137.
- Silver, B. D., B. A. Anderson, and P. R. Abramson. 1986. "Who Overreports Voting? *American Political Science Review* 80(02):613–624.
- Smets, K., and C. van Ham. 2013. "The Embarrassment of Riches? A Meta-analysis of Individual-level Research on Voter Turnout." *Electoral Studies* 32(2):344–359.
- Swaddle, K., and A. Heath. 1989. "Official and Reported Turnout in the British General Election of 1987." *British Journal of Political Science* 19(04):537–551.
- Tingsten, H. 1937. *Political Behavior: Studies in Election Statistics*. London: P.S. King and Son.
- Traugott, M. W., and J. P. Katosh. 1979. "Response Validity in Surveys of Voting Behavior." *Public Opinion Quarterly* 43(3):359–377.
- Wolfinger, R. E., and S. J. Rosenstone. 1980. *Who Votes?* New Haven: Yale University Press.